(3) The bases in the coding parts of a nucleotide sequence shall be listed as triplets (codons). The amino acids corresponding to the codons in the coding parts of a nucleotide sequence shall be typed immediately below the corresponding codons. Where a codon spans an intron, the amino acid symbol shall be typed below the portion of the codon containing two nucleotides.

(4) A nucleotide sequence shall be listed with a maximum of 16 codons or 60 bases per line, with a space provided between each codon or group of 10 bases.

(5) A nucleotide sequence shall be presented, only by a single strand, in the 5 to 3 direction, from left to right.

(6) The enumeration of nucleotide bases shall start at the first base of the sequence with number 1. The enumeration shall be continuous through the whole sequence in the direction 5 to 3. The enumeration shall be marked in the right margin, next to the line containing the one-letter codes for the bases, and giving the number of the last base of that line.

(7) For those nucleotide sequences that are circular in configuration, the enumeration method set forth in paragraph (c)(6) of this section remains applicable with the exception that the designation of the first base of the nucleotide sequence may be made at the option of the applicant.

(d) *Representation of amino acids.* (1) The amino acids in a protein or peptide sequence shall be listed using the three-letter abbreviation with the first letter as an upper case character, as in WIPO Standard ST.25 (1998), Appendix 2, Table 3.

(2) A protein or peptide sequence shall be listed with a maximum of 16 amino acids per line, with a space provided between each amino acid.

(3) An amino acid sequence shall be presented in the amino to carboxy direction, from left to right, and the amino and carboxy groups shall not be presented in the sequence.

(4) The enumeration of amino acids may start at the first amino acid of the first mature protein, with the number 1. When presented, the amino acids preceding the mature protein, *e.g.*, pre-sequences, pro-sequences, pre-pro-sequences and signal sequences, shall have negative numbers, counting backwards starting with the amino acid next to number 1. Otherwise, the enumeration of amino acids shall start at the first amino acid at the amino terminal as number 1. It shall be marked below the sequence every 5 amino acids. The enumeration method for amino acid sequences that is set forth in this section remains applicable for amino acid sequences that are circular in configuration, with the exception that the designation of the first amino acid of the sequence may be made at the option of the applicant.

(5) An amino acid sequence that contains internal terminator symbols (*e.g.*, ''Ter'', ''*'', or ''.'', etc.) may not be represented as a single amino acid sequence, but shall be presented as separate amino acid sequences.

(e) A sequence with a gap or gaps shall be presented as a plurality of separate sequences, with separate sequence identifiers, with the number of separate sequences being equal in number to the number of continuous strings of sequence data. A sequence that is made up of one or more noncontiguous segments of a larger sequence or segments from different sequences shall be presented as a separate sequence.

[63 FR 29635, June 1, 1998, as amended at 69 FR 18803, Apr. 9, 2004; 70 FR 10489, Mar. 4, 2005]

§ 1.823 Requirements for nucleotide and/or amino acid sequences as part of the application.

(a)(1) If the ''Sequence Listing'' required by §1.821(c) is submitted on paper: The ''Sequence Listing,'' setting forth the nucleotide and/or amino acid sequence and associated information in accordance with paragraph (b) of this section, must begin on a new page and must be titled ''Sequence Listing.'' The pages of the ''Sequence Listing'' preferably should be numbered independently of the numbering of the remainder of the application. Each page of the ''Sequence Listing'' shall contain no more than 66 lines and each line shall contain no more than 72 characters. The sheet or sheets presenting a sequence listing may not include material other than part of the sequence listing. A fixed-width font should be

used exclusively throughout the ''Sequence Listing.''

(2) If the ''Sequence Listing'' required by §1.821(c) is submitted on compact disc: The ''Sequence Listing'' must be submitted on a compact disc in compliance with §1.52(e). The compact disc may also contain table information if the application contains table information that may be submitted on a compact disc (§1.52(e)(1)(iii)). The specification must contain an incorporation-by-reference of the Sequence Listing as required by §1.52(e)(5). The presentation of the ''Sequence Listing'' and other materials on compact disc under §1.821(c) does not substitute for the Computer Readable Form that must be submitted on disk, compact disc, or tape in accordance with §1.824.

(b) The ''Sequence Listing'' shall, except as otherwise indicated, include the actual nucleotide and/or amino acid sequence, the numeric identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the ''Sequence Listing.'' The order and presentation of the items of information in the ''Sequence Listing'' shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an ''M'' is mandatory. The submission of those items of information designated with an ''O'' is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the ''Sequence Listing.'' The following table illustrates the numeric identifiers.

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/or Initials.	M.
<120>	Title of Invention		M.
<130>	File Reference	Personal file reference	M when filed prior to assignment of appl. number.
<140>	Current Application Number.	Specify as: US 07/999,999 or PCT/US96/99999.	M, if available.
<141>	Current Filing Date	Specify as: yyyy-mm-dd	M, if available.
<150>	Prior Application Number.	Specify as: US 07/999,999 or PCT/US96/99999.	M, if applicable include priority documents under 35 USC 119 and 120.
<151>	Prior Application Filing Date.	Specify as: yyyy-mm-dd	M, if applicable.
<160>	Number of SEQ ID NOs.	Count includes total number of SEQ ID NOs.	M.
<170>	Software	Name of software used to create the Sequence Listing.	O.
<210>	SEQ ID NO:#:	Response shall be an integer representing the SEQ ID NO shown.	M.
<211>	Length	Respond with an integer expressing the number of bases or amino acid residues.	M.
<212>	Type	Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be "DNA." In addition, the combined DNA/RNA molecule shall be further described in the <220> to <223> feature section.	M.
<213>	Organism	Scientific name, i.e. Genus/ species, Unknown or Artificial Sequence. In addition, the "Unknown" or "Artificial Sequence" organisms shall be further described in the <220> to <223> feature section.	M
<220>	Feature	Leave blank after <220>. <221–223> provide for a description of points of biological significance in the sequence..	M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA"

Numeric identifier	Definition	Comments and format	Mandatory (M) or optional (O).
<221>	Name/Key	Provide appropriate identifier for feature, preferably from WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6.	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence.
<222>	Location	Specify location within sequence; where appropriate state number of first and last bases/amino acids in feature.	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence.
<223>	Other Information	Other relevant information; four lines maximum.	M, under the following conditions: if ''n,'' ''Xaa,'' or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is ''Artificial Sequence'' or ''Unknown''; if molecule is combined DNA/RNA.
<300>	Publication Information.	Leave blank after <300>	O.
<301>	Authors	Preferably max of ten named authors of publication; specify one name per line; preferable format: Surname, Other Names and/or Initials.	O.
<302>	Title		O.
<303>	Journal		O.
<304>	Volume		O.
<305>	Issue		O.
<306>	Pages		O.
<307>	Date	Journal date on which data published; specify as yyyy-mm-dd, MMM-yyyy or Season-yyyy.	O.
<308>	Database Accession Number.	Accession number assigned by database including database name.	O.
<309>	Database Entry Date.	Date of entry in database; specify as yyyy-mm-dd or MMM-yyyy.	O.
<310>	Patent Document Number.	Document number; for patent-type citations only. Specify as, for example, US 07/999,999.	O.
<311>	Patent Filing Date	Document filing date, for patent-type citations only; specify as yyyy-mm-dd.	O.
<312>	Publication Date	Document publication date, for patent-type citations only; specify as yyyy-mm-dd.	O.
<313>	Relevant Residues	FROM (position) TO (position)	O.
<400>	Sequence	SEQ ID NO should follow the numeric identifier and should appear on the line preceding the actual sequence.	M.

[63 FR 29636, June 1, 1998, as amended at 65 FR 54681, Sept. 8, 2000; 68 FR 38630, June 30, 2003]

§ 1.824 Form and format for nucleotide and/or amino acid sequence submissions in computer readable form.

(a) The computer readable form required by §1.821(e) shall meet the following requirements:

(1) The computer readable form shall contain a single ''Sequence Listing'' as either a diskette, series of diskettes, or other permissible media outlined in paragraph (c) of this section.

(2) The ''Sequence Listing'' in paragraph (a)(l) of this section shall be submitted in American Standard Code for Information Interchange (ASCII) text. No other formats shall be allowed.

(3) The computer readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors' or other custom computer programs; however, it shall conform to all requirements detailed in this section.

(4) File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on one of the systems described in paragraph (b) of this section.

(5) Page numbering must not appear within the computer readable form version of the ''Sequence Listing'' file.

(6) All computer readable forms must have a label permanently affixed thereto on which has been hand-printed or typed: the name of the applicant, the title of the invention, the date on